Unsupervised Automatic Defect Inspection based on Image Matching and Local One-class Classification

Chengkan Lv[1,2,3] Zhengtao Zhang[1,2,3] Fei Shen[1,2,3] Feng Zhang[1,2,3]
[1]Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2]The School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
[3]CASI Vision Technology CO., LTD., Luoyang, China
{chengkan.lv, zhengtao.zhang, fei.shen, feng.zhang}@ia.ac.cn

Abstract

In this paper, an unsupervised defect inspection method based on anomaly detection is proposed to inspect various kinds of surface defects in the field of industrial production. This method consists of two modules: (i) An image matching module is utilized to align the input image with a pre-specified template image. Specifically, all objects to be detected will be adjusted to the same position and angle. The aligned images can reduce the difficulty of the training stage, facilitating the subsequent feature extraction and anomaly localization. (ii) After the image matching procedure, an anomaly localization module is trained to learn a mapping that concentrates normal samples in feature space. In particular, each local image region is assigned a feature center by adopting a feature map as the mapping target. Therefore, the compactness of the features extracted from the same region can be improved, which is beneficial to detect potential anomalous targets. Moreover, various artificial defective images are synthesized during the training stage to further improve the discriminatory ability of the anomaly localization module. A series of experiments are conducted on MAD dataset and the industrial production line. The experimental results verify the efficiency and versatility of the proposed method.

1. Introduction

Surface defect inspection is vital to control the product quality in the industrial production line[1, 2]. Traditionally, the inspection model is obtained by supervised learning[3, 4], which requires plenty of labeled defective images to train the model. However, collecting sufficient defective images could be time-consuming in the real production environment. Furthermore, the image annotation also requires a lot of time and labor costs. Therefore, the unsupervised-learning-based anomaly detection method has received increasing attention in both academia and industry[5]. Without requiring actual defective samples during the training stage, anomaly detection focuses on detecting inconsistencies with normal data to locate a

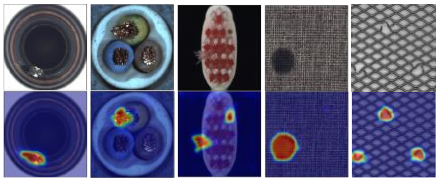


Figure 1: Example images for five products in MAD dataset[14]. For each category, the top row shows the defective images, while the bottom row highlighted the defective regions.

detective regions as shown in Figure 1. This characteristic of anomaly detection can accelerate the adaptation process in a new environment and reduce the production cost.

The anomaly detection methods can be roughly classified into two categories: reconstruction-based method and feature-based method.

The reconstruction-based method aims to obtain a defect-free image by eliminating the anomalous regions. Afterward, the anomalies can be located by analyzing the differences between the original and the reconstructed images[6-11]. Furthermore, References [12, 13] had proved that the segmentation network can be utilized as the discriminative module to locate the defects. However, the average location precision of the reconstruction-based method needs further improvement.

The feature-based method is another popular strategy in anomaly detection. Generally, the distribution of the extracted features from normal images is more regular than that of anomalous images. Therefore, the anomalies can be detected by comparing the differences between input features and normal features.

There are several trends in feature-based anomaly detection methods: Reference [15] introduced knowledge distillation into the anomaly detection method. Since the student networks may learn the representation of normal images from different perspectives under the guidance of the teacher network, the extracted features of student networks may vary on anomalous images. Therefore, the variance of the features of student networks can be utilized to locate anomalous regions. References [16, 17] followed this method and adopted the multiscale strategy to further improve the location accuracy.

Furthermore, another feature-based method is one-class

classification(OCC), which often constructs a discriminative hyperplane to detect anomalies[18]. Specifically, a feature extraction network is trained to map all defect-free images into a pre-specified region in the feature space. Therefore, the features of the anomalous images may be mapped away from the normal features. Hence, a hyperplane surrounding the normal features can be constructed to detect anomalies[19]. With a pre-trained feature extraction network, Reference [20] detected the anomalous regions by calculating the distances between the features of input images and normal images. Reference [21] separately constructed several Gaussian models from multiple feature maps for anomaly detection. Reference [22] adopted two pre-trained networks and designed a mapping method to connect the feature space of the two networks. Afterward, the anomalies can be located since the features of abnormal regions cannot be transmitted smoothly between the two feature spaces. Reference [23] utilized the distance between the feature of the input image and the nearest normal image to locate the defects. Reference [24] improved the discriminatory ability of the feature extraction network by introducing an assistant classification task. In [25], the original one-class dataset was transformed into a multiclass dataset by geometric transformation. Afterward, triplet loss was utilized to emphasize the differentiability between the features of different classes, improving the performance of feature extraction.

However, the following problems remain unsettled.

1) The traditional anomaly detection methods generally fail to consider the situation where significant differences exist in positions and angles between the objects to be detected, leading to performance degradation in this complex scene.

2) The available methods may be unable to balance the accuracy and efficiency. The traditional methods with one global feature center are fast but lack accuracy[26]. To improve the detection performance, some works choose the closest normal feature as the mapping target[23][19]. However, it may significantly affect the efficiency of the method.

Aiming at the above problems, a novel method based on OCC is proposed to detect complex defects accurately and efficiently. The contributions of this paper are summarized as follows:

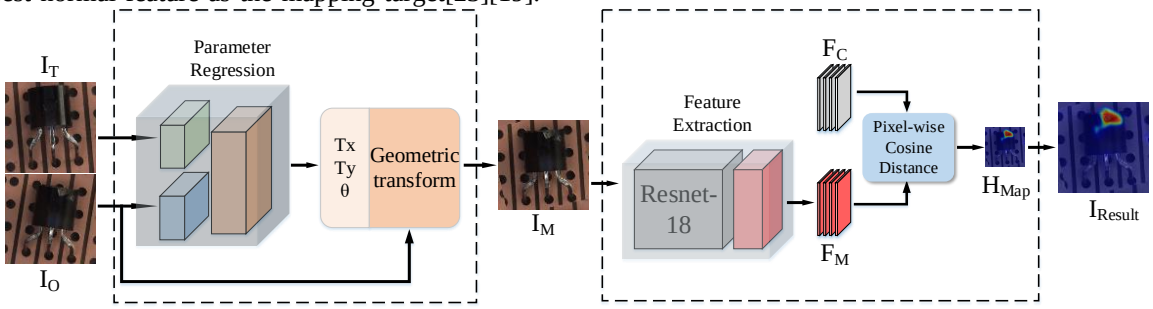
1) To adapt to the scene where significant differences exist in positions and angles between the objects, an image matching module is innovatively introduced into the training phase of anomaly detection to reduce the difficulty of feature extraction process.

2) A detection module based on local OCC is designed to learn compact feature representations improve the accuracy of anomaly localization. Furthermore, a defect generation module is designed to synthesize artificial defective images to improve discriminatory ability of the module.

The remainder of the paper is arranged as follows. Firstly, the proposed defect inspection method will be introduced in detail in Section 2. Secondly, a series of related experiments are conducted to evaluate the performance of our method in Section 3 and Section 4. Finally, conclusions are presented in Section 5.

2. Proposed Method

To deal with the problems mentioned before, an anomaly detection method is proposed to detect defects in complex scenes efficiently. As shown in Figure 2, the proposed method consists of two modules: an image matching module (IMM) and an anomaly localization module (ALM). Firstly, the original image I_O is aligned with the pre-specified template image I_T by the IMM to obtain the aligned image I_M. Specifically, the spatial misalignment including shift and rotation between I_T and I_O is corrected by geometric transformation. After obtaining the aligned image I_M, the feature representations of each local image region of I_M are extracted in ALM, obtaining the feature map F_M.



(a) Image Matching Module (b) Anomaly Localization Module

Figure 2: The overall schema of the proposed method. (a) The procedure of the image matching module. I_T and I_O separately represent the template image and the original image to be detected. Tx, Ty and θ are three geometric transformation parameters obtained by the parameter regression phase. I_M is the aligned image after matching. (b) The procedure of the anomaly localization module. F_M is the representation obtained by the feature extraction phase. F_C is a pre-specified mapping center and H_{Map} is a heatmap which represents the pixel-wise cosine distances between F_M and F_C. I_{Result} is the detection result.

2.1. Image Matching Module

In the industrial production line, significant differences often exist in positions and angles between the objects to be detected. As shown in Figure 3, four defect-free transistors are captured with spatial misalignment. However, the traditional OCC-based methods try to describe these samples with a single model, which may affect the sensitivity to anomalies. As introduced in Section I, the feature extraction network is trained by minimizing the distances from the features of defect-free images to the pre-defined center. To map the diversified objects as shown in Figure 3 into the neighborhood of the same center, the model needs to be robust to the differences in the input images. Consequently, some slight anomalies may be neglected since their small area or the similar structure to the defect-free regions, reducing the accuracy of the method.

Therefore, an image matching module (IMM) is proposed to simulate the mechanical alignment process. Precisely, all the input images will be aligned with the template to improve the consistence of the objects, facilitating the subsequent feature extraction and anomaly localization procedures.

Figure 4 shows the structure of IMM, the template I_T is randomly sampled from the defect-free images using a pre-specified random seed as the target for image matching. For each original image I_O, the geometric transformation parameters are obtained by the sub-networks to align I_O with I_T. Specifically, N_P is a fully convolutional network that extracts the preliminary features of I_T and I_O, obtaining the feature maps F_T and F_D. Afterward, F_T is concatenated with F_D as the input of the regression network N_R, which outputs three geometric transformation parameters: the horizontal translation Tx, the vertical translation Ty, and the rotation angle θ. Finally, according to Tx, Ty and θ, geometric transformation is utilized to align I_O with I_T, obtaining the aligned image I_M. L_{reg} measures the difference between I_M and I_T to train N_P and N_R, which can be formulated as follows:

$$L_{reg}\left(I_M, I_T\right) = \frac{1}{n}\left\|I_M - I_T\right\|_2^2 \tag{1}$$

$$I_M = f_{geo}\left(I_O, N_R\left(N_P\left(I_O\right) \oplus N_P\left(I_T\right)\right)\right) \tag{2}$$

where $\oplus$ represents the concatenation process of two feature maps in the channel dimension. f_{geo} represents the geometric transformation realized by Spatial Transformer Network[27]. n is the number of pixels in the image. Figure 5 shows the results of image matching on real defective images.

As shown in Figure 5, distinct spatial misalignment exists between the objects in the input images I_O. Meanwhile, diverse anomalous regions marked by red circles occur in I_O. However, the objects in the aligned

Figure 3: Images of defect-free objects with spatial misalignment

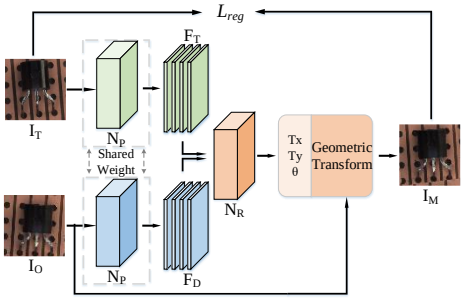


Figure 4: The structure of the image matching module. The two N_P share the same weights and separately extract the features of I_T and I_O. F_T and F_D are the extracted feature maps. N_R is a regression network

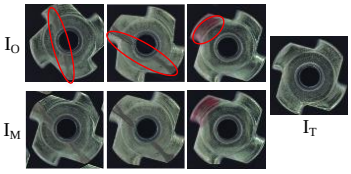


Figure 5: The results of image matching by the trained IMM

image I_M can be basically adjusted into the same position and angle as that in I_T. After the image matching procedure, the consistency of the images can be effectively improved, simplifying the training process of feature extraction and facilitating the followed anomaly localization.

2.2. Basic Anomaly Localization Module based on One-class Classification

Following the image matching procedure, the anomalous regions can be located by the OCC-based module. To balance the accuracy and efficiency of the method, an anomaly localization module (ALM) based on fully convolutional network is utilized to extract compact features of each local image region simultaneously. Afterward, each pixel in the extracted feature map is assigned a local feature center as the mapping target, named as local OCC in this paper.

Subsequently, the pixel-wise cosine distances between FM and the pre-defined mapping center FC are calculated to obtain the heatmap H_{Map}, which indicates the locations of anomalies. Finally, the detection result I_{Result} is gained by up-sampling the low-resolution H_{Map}.

As shown in Figure 6, the ALM consists of two sub-networks, including a pre-trained Resnet-18 and a feature mapping network NM. Firstly, the pre-trained Resnet-18 is utilized for preliminary feature extraction. Secondly, the feature maps from Stage 1, Stage 2 and Stage 3 of Resnet-18 are extracted and merged by scaling and concatenating

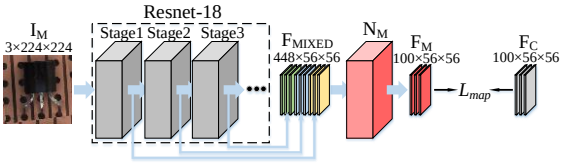


Figure 6: The structure of the anomaly localization module. F_{Mixed} is a multi-scale feature representation of I_M. N_M represents the feature mapping network.

to obtain the multi-scale feature map F_{Mixed}. Thirdly, F_{Mixed} is further mapped by the subsequent mapping network NM to obtain the feature maps F_M for local OCC. Each vector in the F_M is the feature representation of the corresponding local image region in I_M. Afterward, L_{map} measures the pixel-wise cosine distances between F_M and the pre-defined mapping target F_C, where FC is the mean after dimensionality reduction of the F_{Mixed} of all defect-free images. Subsequently, L_{map} is utilized to train the feature mapping network N_M. During this process, the loss function can be formulated as follows:

$$L_{map}\left(F_M, F_C\right) = \frac{1}{wh}\sum_{i=0}^{w-1}\sum_{j=0}^{h-1} 0.5 \times dist\left(F_M^{i,j}, F_C^{i,j}\right) \quad (3)$$

$$dist\left(a,b\right) = 1 - \frac{a \cdot b}{\left\|a\right\|_2 \left\|b\right\|_2} \quad (4)$$

where w and h are the width and height of the feature map F_M. $F_M^{i,j}$ and $F_C^{i,j}$ separately represent the feature vector of F_M and F_C at location (i, j). $dist(a,b)$ is the cosine distance between vector a and vector b, where "$\cdot$" represents the dot product. Under the effect of L_{map}, the network is trained to keep the mapped features $F_M^{i,j}$ in the vicinity of the corresponding center $F_C^{i,j}$. Consequently, in the testing phase, the L_{map} can be utilized to locate the anomalies since the features of anomalous regions are most likely far from the corresponding mapping centers.

Compared with the traditional OCC-based methods that forcedly map all local image regions to the same target, the proposed local OCC-based method adopts the feature map as the mapping target. Therefore, each feature vector in F_M is assigned a different mapping target, which is equivalent to assigning a feature center to each local image regions. Since all the objects in I_M are basically in the same position and angle, the image patches corresponding to the same local center are basically the same, reducing the training difficulty of the method. Consequently, the trained N_M can obtain compact feature representations of defect-free images. Therefore, the method is more sensitive to small differences, which helps to improve the overall detection accuracy.

2.3. Improved Anomaly Localization Module Combined with Outlier Exposure

However, since only defect-free images are utilized in the training stage mentioned above, the performance of the basic ALM on real defective images is limited[28]. Therefore, Outlier Exposure (OE) [29] is introduced in ALM to improve the performance. Specifically, a defect generation module (DGM) is designed to synthesize various artificial defective images to assist in the training of ALM.

As shown in Figure 7, DGM firstly random selects a region on the defect-free images I_M, marked as M. Then, the regions covered by M are randomly darkened or lightened or directly replaced with images from VOC, obtaining an image I_{M+} with anomalous local patterns. Subsequently, the feature map F_M of I_{M+} is extracted by ALM. Afterward, original L_{map} is also modified. Besides the basic task illustrated in Section II.B, the network is further required to maximize the distance from the feature of anomalous region to the mapping center. The modified L_{map} can be formulated as follows:

$$L_{map}\left(F_M, F_C\right) = \frac{1}{wh}\sum_{i=0}^{w}\sum_{j=0}^{h}\Big[\left(1 - M_s^{i,j}\right) \times \left(0.5 \times dist\left(F_M^{i,j}, F_C^{i,j}\right)\right) + M_s^{i,j} \times \left(\max\left(0, \delta - 0.5 \times dist\left(F_M^{i,j}, F_C^{i,j}\right)\right)\right)\Big], M_s^{i,j} \in \{0,1\} \quad (5)$$

where M_s is a downscaled M, whose resolution is consistent with the feature map F_M. $M_s^{i,j}$ represents the value of M_s at location (i, j). $M_s^{i,j} = 0$ represents that the corresponding $F_M^{i,j}$ is extracted from normal region and $M_s^{i,j} = 1$ means $F_M^{i,j}$ is extracted from anomalous region. Therefore, the first part of Formula (5) is consistent with Formula (3) which is utilized to improve the compactness of features from normal regions. The second part of Formula (5) is utilized to maximize the distances from anomalous features to the mapping centers. In particular, to avoid overfitting, an upper limit δ of the distance of anomalous features is designed. When the anomalous feature is far enough to the corresponding center, i.e. $0.5 \times dist(F_M^{i,j}, F_C^{i,j}) > \delta$, the second part of Formula (5) is no longer involved in the calculation of L_{map} under the effect of $\max(0, *)$. δ is set to 0.1 in our experiments. The introduced DGM improves the discriminatory ability of the anomaly detection module. Meanwhile, the diverse local anomalous regions ensure the generalization of the detection network.

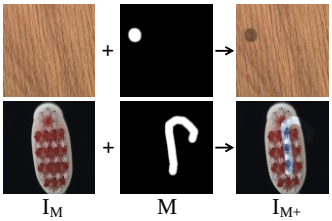


Figure 7: The procedure of DGM.

3. Experiments

To show the effectiveness of the proposed method, several experiments are designed and the experimental

results are analyzed in detail. All the experiments are conducted on a computer equipped with one RTX TITAN GPU, Intel i9-7900X CPU, and an Ubuntu18.04 operating system.

3.1. Experimental setups

Datasets: MAD [14] is a dataset recently proposed for unsupervised anomaly detection mimicking real-world industrial inspection scenarios. MAD consists of fifteen types of images, including five types of texture images and ten types of object images. Furthermore, each category contains dozens of defective images with pixel-level annotations for evaluation. In MAD, the authors have carefully adjusted the object images to ensure that all objects are basically in the same position and angle. To simulate the possible spatial misalignment in the actual production environment, a modified version of the object images in the original MAD is created in this paper, noted as MAD-RTO. Specifically, random rotation and random crop are applied to both the train and test sets of object images.

Implementation details: The proposed IMM and ALM are separately trained and the batch sizes are set to 32 by default. In addition, the learning rates of the two modules are separately set to 2e-4 and 1e-3. Moreover, the training epochs of the two modules are 500 and 300, respectively.

Evaluation metric: To evaluate the performance of the proposed method quantitatively, the area under ROC curve (AUC) is adopted as same as MAD.

3.2. Ablation studies

Firstly, to verify the effectiveness of the proposed basic ALM that utilizes feature map as the mapping center, comparative experiments are conducted on the object images of the original MAD and the modified MAD-RTO.

In Table 1, 'Feature Map' is the proposed method, and 'Vector' represents the method with a global feature vector as the mapping target. Since the objects in the original MAD are basically in the same position and angle, IMM is not utilized in experiments on MAD to avoid redundant operations.

As shown in Table 1, the feature map-based methods outperform the single vector-based methods on MAD and MAD-RTO. Even in the aligned object images, notable discrepancies exist between the local image patches extracted from different regions. Compared with the traditional method with one global center, the proposed ALM assigns a mapping center to each image region. Therefore, the proposed ALM reduces the training difficulty and facilitates the followed anomaly localization, especially in the Transistor, Metal Nut and Cable images with complex structures.

Secondly, to evaluate the effect of the proposed IMM and DGM, six ablative settings are designed on the modified MAD-RTO dataset and the basic ALM is adopted in all these methods. (i) The first method is the proposed method which combines the IMM and the DGM. (ii) The second method removes the IMM, directly localizing the anomalies on the images with spatial misalignment. (iii) The third method is based on the second method with additional data augmentation (AUG), including rotation and translation on the input images. (iv) The fourth method removes the DGM from the first method and utilizes only normal images in the training stages. (v) The fifth method is the basic ALM without IMM and DGM. (vi) The sixth method adds the data augmentation process to the fifth method. Table 2 presents the AUC of each method.

Table 1: The AUC of anomaly localization of the methods with different mapping targets on MAD and MAD-RTO

Object	MAD		MAD-RTO	
	Feature Map	Vector	Feature Map	Vector
Bottle	0.985	0.983	0.990	0.984
Cable	0.978	0.953	0.975	0.945
Capsule	0.985	0.985	0.981	0.971
Hazelnut	0.984	0.984	0.980	0.982
Metal Nut	0.976	0.968	0.972	0.961
Pill	0.971	0.972	0.965	0.970
Screw	0.965	0.970	0.974	0.981
Toothbrush	0.988	0.983	0.982	0.982
Transistor	0.975	0.819	0.978	0.817
Zipper	0.992	0.991	0.992	0.979
Mean	0.980	0.961	0.979	0.957

As shown in column (v,vi) of Table 2, the commonly utilized data augmentation strategies, including rotation and translation, are unable to effectively improve the performance on the MAD-RTO dataset with spatial misalignment. Meanwhile, the same conclusion can be drawn from column (ii,iii) of Table 2. Conversely, as shown in column (iv,v) of Table 2, the proposed IMM significantly improves the performance of anomaly localization. In the MAD-RTO, the image patches extracted from the same local region differ from each other. Consequently, the corresponding feature vectors $F_{Mixed}^{i,j}$ are loosely distributed. Although the network can map all these features to the same centroid after training, the feature of

Table 2: The AUC in anomaly localization experiments with the six methods on MAD-RTO

Objects	Methods					
	(i) IMM + DGM	(ii) DGM	(iii) DGM + AUG	(iv) IMM	(v)	(vi) AUG
Bottle	**0.990**	0.982	0.982	0.962	0.912	0.904
Cable	**0.975**	0.963	0.969	0.933	0.862	0.901
Capsule	**0.981**	0.957	0.958	0.960	0.910	0.922
Hazelnut	0.980	**0.984**	0.985	0.884	0.884	0.903
Metal Nut	**0.972**	0.964	0.967	0.932	0.790	0.787
Pill	**0.965**	0.941	0.937	0.880	0.816	0.821
Screw	**0.974**	0.956	0.962	0.920	0.827	0.822
Toothbrush	**0.982**	0.976	0.974	0.961	0.873	0.858
Transistor	**0.978**	0.928	0.928	**0.978**	0.913	0.912
Zipper	**0.992**	0.991	0.989	0.922	0.910	0.892
Mean	**0.979**	0.964	0.965	0.933	0.870	0.872

small anomalous region may be close to a normal feature due to the loose distribution. In this case, the network will also map these anomalous features near the target center, affecting the performance of the localization. In the image aligned by IMM, the local image regions corresponding to the same feature center are basically the same, i.e., the distribution of the corresponding $F_{Mixed}^{i,j}$ is more compact. Consequently, anomalous features are more likely to fall outside the region of normal features in the testing phase, obtaining better localization performance than the original method.

Furthermore, as shown in column (ii) and column (v) of Table 2, the introduced DGM also effectively improves the accuracy. Although there is a difference between the artificial and the actual defective images, Formula (5) forces the ALM to have sufficient discriminatory ability for the possible anomalous regions. Specifically, the network is prompted to map these generated images to areas away from the center, facilitating the detection of the actual defective images.

Combing the IMM and DGM, the proposed method obtains the best performance on the MAD-RTO dataset, especially in Pill, Capsule and Screw.

3.3. Comparative experiments on the original MAD

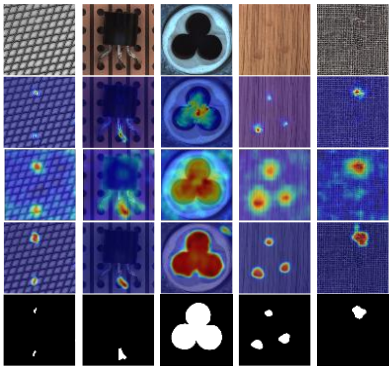


Figure 8: The detection results of the three methods on MAD. The first row shows the anomalous images. The middle three rows separately show the results of STFPM, PADIM and our method. The last row shows the ground truths.

Figure 8 shows the qualitative comparison with our model and two recently proposed methods including PADIM[20] and STFPM[16], where the red regions indicate the anomalies. As can be seen from Figure 8, the proposed method is able to precisely locate the anomalous regions with few misclassified background pixels. Meanwhile, Figure 9 shows the detection results of the proposed method on 15 defective images in MAD.

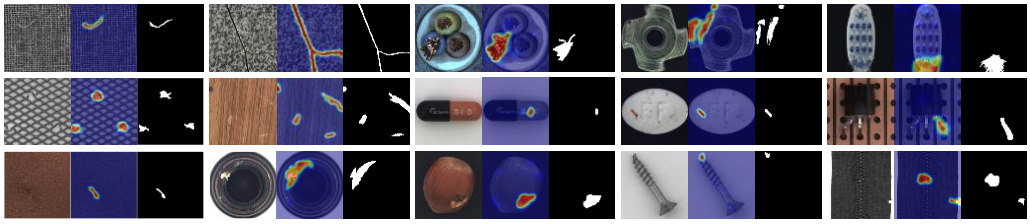


Figure 9: The detection results on 15 images in MAD. The three columns in each group of images separately represent the original defective images, the detection results and the GT.

Table 3: The AUC of anomaly localization of our method and several recently proposed methods on MAD

Class	Ours	PFM	SPADE	PADIM	Patch SVDD	FCDD	STFPM	USTD	MKD	DRAEM	RIAD	AE-SSIM	AE-L2
Carpet	**0.991**	0.992	0.975	**0.991**	0.926	0.96	0.988	0.935	0.956	0.955	0.963	0.87	0.59
Grid	0.991	0.988	0.937	0.973	0.962	0.91	0.990	0.899	0.918	**0.997**	0.988	0.94	0.90
Leather	**0.995**	0.994	0.976	0.992	0.974	0.98	0.993	0.978	0.980	0.986	0.994	0.78	0.75
Tile	**0.974**	0.962	0.874	0.941	0.914	0.91	**0.974**	0.925	0.828	0.992	0.891	0.59	0.51
Wood	0.970	0.962	0.885	0.949	0.908	0.88	**0.972**	0.921	0.848	0.964	0.858	0.73	0.73
Texture Mean	**0.984**	0.979	0.929	0.969	0.937	0.928	0.983	0.931	0.906	0.979	0.939	0.782	0.696
Bottle	0.985	0.984	0.984	0.983	0.981	0.97	0.988	0.978	0.963	**0.991**	0.984	0.93	0.86
Cable	**0.978**	0.967	0.972	0.967	0.968	0.90	0.955	0.919	0.824	0.947	0.842	0.82	0.86
Capsule	0.985	0.983	**0.990**	0.985	0.958	0.93	0.983	0.968	0.958	0.943	0.928	0.94	0.88
Hazelnut	0.984	0.991	0.991	0.982	0.975	0.95	0.985	0.982	0.946	**0.997**	0.961	0.97	0.95
Metal Nut	0.976	0.972	0.981	0.972	0.980	0.94	0.976	0.972	0.864	**0.995**	0.925	0.89	0.86
Pill	0.971	0.972	0.965	0.957	0.951	0.81	**0.978**	0.965	0.896	0.976	0.957	0.91	0.85
Screw	0.965	0.987	**0.989**	0.985	0.957	0.86	0.983	0.974	0.959	0.976	0.988	0.96	0.96
Toothbrush	0.988	0.986	0.979	0.988	0.981	0.94	**0.989**	0.979	0.961	0.981	**0.989**	0.92	0.93
Transistor	**0.975**	0.878	0.941	**0.975**	0.970	0.88	0.825	0.737	0.764	0.909	0.877	0.90	0.86
Zipper	**0.992**	0.982	0.965	0.985	0.951	0.92	0.985	0.956	0.939	0.988	0.978	0.88	0.77
Object Mean	**0.980**	0.970	0.976	0.978	0.967	0.91	0.965	0.943	0.907	0.970	0.943	0.912	0.878
All Mean	**0.981**	0.973	0.960	0.975	0.957	0.916	0.971	0.939	0.907	0.973	0.941	0.868	0.817

To further evaluate the effectiveness of the proposed method, comparative experiments between our method and several recently proposed methods are conducted on MAD, including PFM[22], SPADE[19], PADIM, Patch SVDD[23], FCDD[26], STFPM, USTD[15], DRAEM[12], MKD[17], RIAD[5] and AE[14]. Since the well-aligned images in the original MAD, the proposed IMM is not utilized to avoid redundancy in this stage. Table 3 shows the AUC in pixel-wise anomaly localization and image-level anomaly detection of each method.

The results of USTD are obtained from the reproduction of RIAD, and the results of the remaining methods are obtained from the original papers. The best result in each kind of image is highlighted in boldface.

As shown in Table 3, the proposed method obtains satisfactory performance in both texture and object classes on the well aligned MAD dataset, which can be owed to two factors.

Firstly, the feature map is utilized as the mapping center while training the feature mapping network N_M. Even the objects to be detected are well aligned, the image patches extracted from different region vary from each other. Therefore, when only one global center vector is assigned, the network must adapt to the significant difference between the patches, affecting the sensitivity to actual anomalies. Conversely, the patches acquired from the same local region of different images are similar to each other. Subsequently, when mapping these basically consistent patches to the corresponding local center, the network can pay more attention to minimizing the distances to the target, obtaining compact feature representation and facilitating the followed anomaly localization.

Secondly, the artificial defective images are introduced in the proposed method. Due to the proposed DGM, the mapping network N_M is prompted to map the features of the anomalous regions away from the local center while minimizing the distances of the normal features. This multi-task training strategy strengthens the discriminative ability of the network and optimizes the feature extraction process. Furthermore, the diversified artificial defective images relatively simulate some actual defects such as color change, breakage and crack, promoting the anomaly localization during the testing phase.

3.4. Comparative experiments on the modified MAD-RTO

To evaluate the performance of the proposed method in complex scenarios with misalignment in the position and angle, experiments are also conducted on the modified dataset MAD-RTO. Figure 10 shows the detection results on MAD-RTO. Furthermore, the performance of the proposed method is quantitatively compared with the other methods, including PFM, SPADE, PADIM, Patch SVDD, FCDD, MKD, STFPM and DRAEM. We reproduce these methods in our environment. The AUCs of these methods are shown in Table 4, and the AUCs on the object images in the original MAD are also listed for comparison. However, some related works have pointed out that the AUC does not reflect the localization accuracy well in surface anomaly detection where only a few pixels are anomalous[15]. Therefore, average precision (AP), a more suitable metric for class imbalance scenario like defect inspection, is also adopted for evaluation in Table 4.

Meanwhile, Table 5 presents the average execution time of each method.

As shown in Table 4, when spatial misalignment exists in the images to be tested, performance degradations occur in all traditional methods. Due to the introduction of rotation, the task of identifying relative positions designed in Patch SVDD may not be feasible anymore, affecting the feature extraction process. Moreover, the spatial misalignment in MAD-RTO further increases the looseness of the distribution of image patches and raises the training difficulty of FCDD. As for MKD and STFPM, the teacher network pre-trained on ImageNet has little knowledge about processing input images with rotation. Hence, the feature extraction of the teacher network is more disordered, affecting the learning stage of the student network and the anomaly localization process of these methods. For the reconstruction-based method DRAEM, the involved geometric transformation increased the diversity of input samples, which may affect the inspection of discriminative sub-network.

Although SPADE and PADIM still perform well on MAD-RTO, these are achieved at the cost of real-time performance. As shown in Table 5, the execution times of SPADE and PADIM are longer than that of other methods. Consequently, it may affect the application of these methods in real industrial scenarios. Conversely, the proposed method achieves a good balance between the accuracy and efficiency of detection.

As shown in Table 5, the execution time of our method is kept at the same level as that of PFM and FCDD. Based

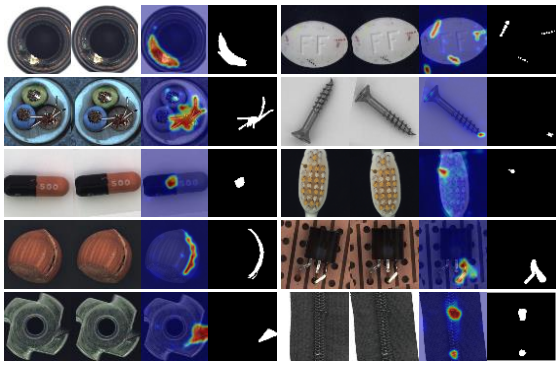


Figure 10: The results of image matching and anomaly detection on 10 object images from MAD-RTO. The four columns in each group separately represent the original images, the aligned images, the detection results and the ground truths.

Table 4: Results for the task of anomaly localization experiments with our method and several recently proposed methods on MAD-RTO (AUC/AP)

	Ours	PFM	SPADE	PADIM	Patch SVDD	FCDD	MKD	STFPM	DRAEM
Bottle	**0.990/0.890**	0.980/0.747	0.959/0.718	0.975/0.725	0.972/0.386	0.958/0.640	0.924/0.529	0.983/0.761	0.955/0.767
Cable	**0.975/0.722**	0.955/0.537	0.899/0.351	0.947/0.376	0.903/0.386	0.909/0.581	0.768/0.148	0.935/0.488	0.854/0.522
Capsule	**0.981/0.487**	0.968/0.390	0.973/0.456	0.970/0.321	0.861/0.127	0.840/0.167	0.942/0.215	0.961/0.308	0.871/0.270
Hazelnut	0.980/**0.738**	**0.988**/0.667	0.982/0.660	0.977/0.553	0.911/0.172	0.948/0.388	0.921/0.297	0.981/0.541	0.846/0.371
Metal Nut	0.972/0.876	0.957/0.771	0.947/0.767	0.946/0.692	0.967/0.837	0.897/0.599	0.791/0.361	0.960/0.765	**0.989/0.954**
Pill	**0.965/0.660**	0.967/0.754	0.946/0.604	0.956/0.559	0.868/0.485	0.813/0.394	0.880/0.264	0.924/0.537	0.954/0.549
Screw	0.974/**0.441**	0.981/0.284	**0.989**/0.403	0.964/0.280	0.963/0.190	0.861/0.031	0.957/0.072	0.976/0.222	0.978/0.295
Toothbrush	0.982/**0.606**	0.984/0.431	**0.985**/0.547	0.976/0.447	0.921/0.181	0.907/0.145	0.945/0.373	**0.985**/0.382	0.957/0.289
Transistor	**0.978/0.804**	0.823/0.468	0.824/0.360	0.938/0.598	0.863/0.482	0.826/0.434	0.681/0.127	0.764/0.361	0.664/0.197
Zipper	**0.992/0.825**	0.974/0.597	0.988/0.753	0.983/0.555	0.903/0.386	0.930/0.291	0.899/0.242	0.968/0.625	0.960/0.651
Mean	**0.979/0.705**	0.958/0.565	0.949/0.562	0.963/0.511	0.913/0.361	0.890/0.367	0.871/0.263	0.944/0.499	0.903/0.486
Mean on MAD	**0.980/0.696**	0.970/0.563	0.976/0.593	0.978/0.576	0.967/0.473	0.910/0.384	0.900/0.279	0.965/0.493	0.970/0.684

Table 5: The execution time of each method

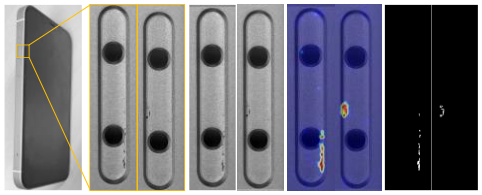
	Ours	PFM	SPADE	PADIM	Patch SVDD	FCDD	MKD	STFPM	DRAEM
Time (ms)	27	14	2625	862	818	13	18	26	32

on the high-performance localization model ALM, a simple and effective matching module IMM is utilized to eliminate the differences existing among the input images in advance. Therefore, the detection accuracy can be maintained without significantly increasing the execution time.

4. Experiments on production line

To further prove the necessity and effectiveness of proposed modules, experiments are also conducted on middle frame (MF) images captured by our inspection equipment in the industrial production line.

As shown in Figure 11, the spatial misalignment exists in the original MF images can be basically repaired in the aligned images. Afterward, the defective regions can be well located by the proposed method with few misclassified background pixels. Table 6 shows the ablation study on MF images. Due to the complex structure of the MF image and the spatial misplacement, both ALM and IMM effectively improve the performance. By introducing artificial anomalous images, DGM further improves the accuracy of the inspection model. Meanwhile, Table 7 shows the quantitative comparison between our method and other methods.



(a) (b) (c) (d) (e)

Figure 11: The detection results on MF dataset. (a) The mobile phone image. (b) Defective circular hole images extracted from the vertical side region of MF. (c) The aligned images. (d) The detection results of our method. (e) The ground truths.

Table 6 The AUC/AP in anomaly localization experiments with the five methods on MF images

	Module			Metric	
	ALM	IMM	DGM	AUC	AP
(i)				0.641	0.013
(ii)	✔			0.855	0.078
(iii)	✔	✔		0.906	0.255
(iv)	✔		✔	0.968	0.559
(v)	✔	✔	✔	0.987	0.680

Table 7: Results of anomaly detection of our method and several recently proposed methods on MF dataset (AUC/AP)

	Ours	PADIM	PFM	SPADE	STFPM	FCDD	MKD
AUC	**0.987**	**0.987**	0.985	0.924	0.823	0.940	0.960
AP	**0.680**	0.556	0.474	0.129	0.064	0.167	0.250

5. Conclusions

In this paper, a method based on OCC is proposed to perform precise and efficient defect inspection without requiring actual defective images. To improves the consistency of the input images, an image matching module is designed to align the input images to facilitate the feature extraction and anomaly localization processes. Afterward, a local OCC-based anomaly localization module is conducted to obtain compact representations of defect-free images. Furthermore, a defect generation module is designed to improve the discriminatory ability of the anomaly localization module. Experimental results show that, without significantly increasing the execution time, the AUROC of the proposed method on MAD and MAD-RTO are 0.981 and 0.979, separately. Meanwhile, it also obtains satisfactory performance on the images extracted from the industrial production line. In future, this approach can be extended to detect tiny defects on images with high resolution and complex structure.

References

[1] Wendong Ding, Zhengtao Zhang, Dapeng Zhang, De Xu, Haibing Lv, Xinxiang Miao, Guorui Zhou and Hao Liu. An effective on-line surface particles inspection instrument for large aperture optical element. *International Journal of Automation and Computing*, 14(4):420-431, 2017.

[2] Liwei Han and De Xu. Statistic learning-based defect detection for twill fabrics. *International Journal of Automation and Computing*, 7(1):86-94, 2010.

[3] Binyi Su, Haiyong Chen, Peng Chen, Guibin Bian, Kun Liu, and Weipeng Liu. Deep learning-based solar-cell manufacturing defect detection with complementary attention network. *IEEE Transactions on Industrial Informatics*, 17(6):4084-4095, 2021.

[4] Jiaxiang Luo, Zhiyu Yang, Shipeng Li, and Yilin Wu. FPCB Surface Defect detection: a decoupled two-stage object detection framework. *IEEE Transactions on Instrumentation and Measurement*, 70: 1-11, 2021.

[5] Zavrtanik Vitjan, Matej Kristan, and Danijel Skočaj. Reconstruction by inpainting for visual anomaly detection. *Pattern Recognition*, 112: 1-9, 2021.

[6] Bergmann Paul, Sindy Löwe, Michael Fauser, David Sattlegger, and Carsten Steger. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. *arXiv preprint arXiv:1807.02011*, 2018.

[7] Perera Pramuditha, Ramesh Nallapati, and Bing Xiang. Ocgan: One-class novelty detection using gans with constrained latent representations. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2898-2906, 2019.

[8] Tian Hu, and Fei Li. Autoencoder-based fabric defect detection with cross-patch similarity. *In 2019 16th International Conference on Machine Vision Applications (MVA)*, pages 1-6, 2019.

[9] Gong Dong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. *In Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1705-1714, 2019.

[10] Schlegl Thomas, Philipp Seeböck, Sebastian M. Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Medical image analysis*, 50: 30-44, 2019.

[11] Akcay Samet, Amir Atapour-Abarghouei, and Toby P. Breckon. Ganomaly: semisupervised anomaly detection via adversarial training. *In Proceedings of Asian Conference Computer Vision*, pages 622-637, 2018.

[12] Zavrtanik Vitjan, Matej Kristan, and Danijel Skočaj. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 8330-8339, 2021.

[13] Chengkan Lv, Fei Shen, Zhengtao Zhang, De Xu and Yonghao He. A novel pixel-wise defect inspection method based on stable background reconstruction. *IEEE Transactions on Instrumentation and Measurement*, 70: 1-13, 2020.

[14] Bergmann Paul, Michael Fauser, David Sattlegger, and Carsten Steger. MVTec AD -- A comprehensive real-world dataset for unsupervised anomaly detection. In Proceedings of *IEEE Conference on Computer Vision and Pattern Recognition*, pages 9592-9600, 2019.

[15] Bergmann Paul, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: student-teacher anomaly detection with discriminative latent embeddings. *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 4182-4191, 2020.

[16] Guodong Wang, Shumin Han, Errui Ding and Di Huang. Student-teacher feature pyramid matching for unsupervised anomaly detection. *arXiv preprint arXiv:2103.04257*, 2022.

[17] Salehi Mohammadreza, Niousha Sadjadi, Soroosh Baselizadeh, Mohammad H. Rohban, and Hamid R. Rabiee. Multiresolution knowledge distillation for anomaly detection. *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 14902-14912, 2021.

[18] Ruff Lukas, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. *In Proceedings of International Conference on Machine Learning*, pages 6981-6996, 2018.

[19] Cohen Niv, and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. *arXiv preprint arXiv:2005.02357*, 2022.

[20] Defard Thomas, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Audigier. PaDiM: a patch distribution modeling framework for anomaly detection and localization. *In Proceedings of International Conference on Pattern Recognition*, pages 475-489, 2021.

[21] Rippel Oliver, Patrick Mertens, Eike König, and Dorit Merhof. Gaussian anomaly detection by modeling the distribution of normal data in pretrained deep features. *IEEE Transactions on Instrumentation and Measurement*, 70: 1-13, 2021.

[22] Wan Qian, Liang Gao, Xinyu Li, and Long Wen Unsupervised image anomaly detection and segmentation based on pre-trained feature mapping. *IEEE Transactions on Industrial Informatics*, 2022.

[23] Jihun Yi, and Sungroh Yoon. Patch SVDD: patch-level svdd for anomaly detection and segmentation. *In Proceedings of Asian Conference Computer Vision*, pages 375-390, 2020.

[24] Perera Pramuditha, and Vishal M. Patel. Learning deep features for one-class classification. *IEEE Transactions Image Processing*, 28(11):5450-5463, 2019.

[25] Bergman Liron, and Yedid Hoshen. Classification-based anomaly detection for general data. *In Proceedings of International Conference of Learning Representations*, 2020.

[26] Liznerski Philipp, Lukas Ruff, Robert A. Vandermeulen, Billy Joe Franks, Marius Kloft, and Klaus-Robert Müller. Explainable deep one-class classification. *In Proceedings of International Conference of Learning Representations*, 2021.

[27] Jaderberg, Max, Karen Simonyan, and Andrew Zisserman. Spatial transformer networks. *In Proceedings of Annual Conference on Neural Information Processing Systems*, pages 2017-2025, 2015.

[28] Ruff Lukas, Robert A. Vandermeulen, Nico Görnitz, Alexander Binder, Emmanuel Müller, Klaus-Robert Müller, and Marius Kloft. Deep semi-supervised anomaly detection," *In Proceedings of International Conference of Learning Representations*, 2020.

[29] Hendrycks Dan, Mantas Mazeika, and Thomas Dietterich. Deep anomaly detection with outlier exposure. *In Proceedings of International Conference of Learning Representations*, 2019.